IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 26, 2019, accepted February 19, 2019, date of publication February 28, 2019, date of current version April 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2902243

An Energy Aware Adaptive Kernel Density Estimation Approach to Unequal Clustering in Wireless Sensor Networks

FAGUI LIU AND YUFEI CHANG
School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

Corresponding author: Yufei Chang (201720139968@mail.scut.edu.cn)

This work was supported in part by the Engineering and Technology Research Center of Guangdong Province for Logistics Supply Chain and Internet of Things under Project GDDST[2016]176, in part by the Key Laboratory of Cloud Computing for Super—Integration Cloud Computing in Guangdong Province under Project 610245048129, in part by the Engineering and Technology Research Center of Guangdong Province for Big Data Intelligent Processing under Project GDDST[2013]1513-1-11.

ABSTRACT Energy conservation is one of the most important challenges in wireless sensor networks (WSNs). Therefore, compared with the traditional networks, the WSNs not only need high-quality services with high throughput or low transmission delay, but also pay greater attention to energy utilization to extend network lifetime. The clustering routing algorithm is considered to be among the effective ways to collect and transmit data in WSNs. Cluster head (CH) plays a vital role in the cluster which is in charge of data aggregation and data transmission, so their energy consumption is higher than non-CH nodes. The traditional clustering algorithm tends to have the same size in each cluster. However, due to the randomness of the node distribution, the equal clustering mechanism obviously cannot reduce energy consumption. In order to solve this problem, this paper contributes a new unequal clustering algorithm, an energy-aware adaptive kernel density estimation algorithm (EAKDE), which aims to balance the energy dissipation among the CHs. EAKDE utilizes fuzzy logic to determine the priority of nodes competing for CH. In order to adapt the dynamic change of node conditions, adaptive kernel density estimation algorithm is utilized to assign the appropriate unequal cluster radius to sensor nodes. The simulation results demonstrate that, in different scenarios, EAKDE outperforms the other well-known algorithms in terms of network stability, network lifetime, and energy efficiency.

INDEX TERMS Unequal clustering, fuzzy logic, kernel density estimation, wireless sensor networks.

I. INTRODUCTION

With the development of micro-electro-mechanical systems (MEMS) technology [1], wireless communications and Internet of Things (IoT), there are many applications for wireless sensor networks (WSNs), such as traffic control, smart city, environmental monitoring, health care and disaster area monitoring [2]. In order to detect the environment factors (e.g., temperature, moisture, pressure and electromagnetic environment) for the region of interest, a large number of sensor nodes are deployed in the field. Sensor nodes report to the base station (BS) when an event is detected. The BS is a gateway between the end user and the sensor node, where the user can acquire the related information from BS through the Internet.

The associate editor coordinating the review of this manuscript and approving it for publication was Jing Liang.

Typically, sensor nodes are cheap and small-sized devices with small memory; hence, their energy supply and processing power are very limited. In addition, sensor nodes are usually dispersed in extremely harsh environments that are access-limited to humans, which makes it impractical to replace the node battery. Therefore, the energy efficiency of sensor nodes is critical to the network lifetime. Node energy is primarily consumed in environmental sensing, data processing, and data transmission. Compared to data transmission, other overhead is relatively little. That is what makes the choice of routing protocols extremely critical. In the literature [40], [41], many clustering routing algorithms have been proposed.

In order to overcome the huge energy consumption for direct transmission to BS, cluster-based transmission protocols are widely researched and applied [3]. The WSNs is divided into a group of clusters, each of which has a coordinator called the cluster head (CH). The sensing data collected by cluster member (CM) nodes are not directly sent to BS, but the corresponding CH. The responsibilities of the CHs are aggregating data from CMs and forwarding it to BS.

Some clustering routing algorithms are proposed in the literature. These methods aim at reducing energy consumption in WSNs and improving the entire network lifetime. LEACH [4] employs random CH selection mechanism and periodically rotates the role of CHs to balance energy consumption. EAMR [20] forms clusters in a LEACH-like manner, with the difference that EAMR utilizes fixed clustering, multi-hop routing, and threshold-based CH selection mechanism. The improved K-means routing scheme (imp-K-means) [19], based on the K-means, adopts the equal clustering mechanism to set up clusters. EAUCF [31] is based on a purely probabilistic model to select tentative CHs, and takes advantage of fuzzy logic to allocate competition radius. DFCR [32] is proposed to solve the hot spot problem. It takes the energy level, neighbor density, intra-cluster transmission cost and distance to BS, as the reference basis for CH selection and cluster radius calculation.

In this paper, we propose an energy aware adaptive kernel density estimation algorithm (EAKDE) to unequal clustering. Different from the traditional random CH selection method, EAKDE considers the residual energy and distance to BS, and employs fuzzy-logic-based approach to determine the priority of a node to compete for CH. Then, according to the local node information, EAKDE adaptively decides the cluster radius. After the cluster is formed, EAKDE utilizes a multi-hop routing protocol. The CH transmits data between the clusters through a relay node, and finally sends it to BS. These mechanisms significantly reduce the energy overhead of sensor nodes, so the network lifetime is ultimately increased.

To evaluate the proposed algorithm, EAKDE will be compared to some popular algorithms proposed in the literature, namely LEACH, EAMR, imp-K-means, EAUCF, and DFCR. Simulation experiments are performed on two different scenarios. The simulation results show that EAKDE performs better than the other advanced algorithms in terms of network stability, network lifetime and energy efficiency.

The rest of this paper is organized as follows. In the next section, related research on some clustering routing algorithms in WSNs is given briefly. In Section 3, the network model and the energy consumption model used in this paper are introduced. In Section 4, the proposed algorithm EAKDE and its four main phases are discussed in detail. In Section 5, in order to evaluate EAKDE, it is compared with LEACH, EAMR, imp-K-means, EAUCF and DFCR, and the detailed experimental results are given. Finally, the conclusions and future works are drawn in Section 6.

II. RELATED WORK

Sensing data transmission is deemed to be the most important energy consumption of sensor nodes in WSNs. Clustering routing protocols not only extend the network lifetime, but also increase the network scalability. In this section, many clustering routing algorithms proposed in recent years are briefly explained.

LEACH [4] is a well-known distributed clustering algorithm. LEACH operates on rounds where each round divided into two phases: the setup phase and the steady state phase. During the setup phase, LEACH uses a pure probability model to select CHs. Each sensor node has a certain probability of acting as a CH per round. With attention on energy dissipation, the CHs take a rotation among all sensor nodes in each round. When a node is decided to be CH, it broadcasts an advertisement message, with the node id and a header. Non-CHs decide to join the cluster according to the strength of received signal. In the steady state phase, the CHs aggregate data packets received from their CMs and forward them to BS directly.

Based on LEACH, many variants have been proposed to further improve its performance. The LEACH-based clustering protocols mainly extend the network lifetime from the perspective of CH selection and cluster formation. The improved LEACH (ModLEACH) [7] proposes a CH replacement scheme. When the CH's energy is below a given threshold, the CH is re-elected, otherwise, it will continue to act as CH. LEACH-C [5] adopts a centralized mechanism to select CHs based on the residual energy of each node and location awareness information. LEACH-E [6] uses minimum spanning tree technology to select CHs based on residual energy of each node. In addition, some protocols consider combining heuristic-based algorithms, namely PSO [10], GA [8], and ACO [10], [12]. For example, LEACH-GA [8] employs a genetic algorithm to select CHs with optimal probability. HAS [13] adopts a centralized protocol, which aims to minimize the total distance between the CH and its CMs, thereby further enhancing the network lifetime.

After given the number of CHs, K-means [16]–[18] algorithm usually considers the Euclidean distance to determine the centroid position, and the node closest to the centroid serves as CH. Imp-K-means [19] is divided into two phases. The first phase is similar to K-means, except that imp-K-means considers the residual energy of each node to optimize CH selection. The next stage of imp-K-means utilizes equal clustering schemes to ensure that the clusters have the same size. This approach sounds good for the uniform distribution of nodes. However, in most scenarios, the node location distribution and energy distribution are often uneven, so the equal clustering mechanism is obviously not applicable to most of the WSNs applications. To solve this problem, the unequal clustering mechanism is proposed. EEUC [9] utilizes a probabilistic model to determine whether a sensor node participates in CH selection. If a sensor node takes part in the campaign, it will act as a temporary CH and participate in the campaign within a pre-specified competitive radius to become an actual CH. The competition radius is proportional to its distance to BS. The sensor node near BS has a smaller competition radius. Therefore, EEUC is a distributed unequal clustering algorithm.

Multi-hop routing protocols in WSNs tend to consume less power than traditional single-hop communication protocols. In addition, multi-hop routing protocols can effectively overcome signal propagation effects in remote wireless communications, and improve communication quality of service (QoS). EAMR is an energy-efficient multi-hop routing protocol. EAMR minimizes the communication overhead for exchanging control information by reducing the number of CH rotations. When the residual energy of a CH is less than a given threshold, another node is randomly selected from its CMs to serve as CH. Therefore, the CMs are fixed throughout the process. In addition, EAMR uses a relay node (RN) for data packet transmission between clusters. In each round, the CH far from BS transmits data to BS through RN, and the CH close to BS directly transmits to BS.

The research application of fuzzy logic [23]–[28] in WSNs is currently more popular. FCM was proposed by Bezdek in 1981 [21], which allows sensor nodes to belong to multiple clusters. Each node has a membership degree of each cluster in the interval [0, 1] [22]. This method divides all nodes in WSNs into a given number of clusters. In addition, many algorithms utilize fuzzy logic to solve CH selection and cluster radius calculation problems. CHEF [29] uses the residual energy of each node and local distance as fuzzy input descriptors to select CHs in a distributed manner. The local distance refers to the distance between the temporary CH and the CMs within its competition radius. EAUCF [31] utilizes fuzzy logic to deal with the uncertainties in the competition radius and solves the hot spot problem in multi-hop routing. EAUCF uses the node residual energy and distance to BS to adjust the competition radius, so the competition radius of each round is dynamically changed.

DUCF [30] and DFCR [32] which are based on EAUCF, consider more fuzzy input parameters and mapping rules. DUCF takes the residual energy of each node, node degree and distance to BS as fuzzy input parameters. The fuzzy output result serves as a reference for CH selection and competition radius calculation. Node degree is the number of neighbor nodes within a given communication radius. However, DFCR considers four fuzzy input parameters, including the primary parameters: the energy level and distance to BS, and the secondary parameters: neighbor density and neighbor cost. The CH selection is determined by the primary parameters, while the cluster radius is calculated according to the primary parameters and the secondary parameters. This approach effectively enhances the network lifetime. However, more fuzzy input parameters and if-then mapping rules are adopted, which makes DFCR greatly influenced by human experience.

The proposed algorithm EAKDE in this paper is mainly divided into four phases:

- Cluster head election: Using fuzzy logic theory, EAKDE calculates the priority of each node competing for CH based on the node residual energy and distance to BS.
- Cluster radius calculation: According to the local sensor nodes information, the adaptive kernel density estimation algorithm is adopted to determine the cluster radius of sensor nodes.
- Cluster formation: Each CH sends a broadcast message within its cluster radius calculated in step 2. For a non-CH node that receives multiple messages from CHs, it considers four parameters, (i.e., distance of CH from non-CH, residual energy of CH, direction of CH to BS, and distance of CH from BS) and decides to join the optimal cluster.
- Routing process: The CH node level is measured by the distance of CH from BS. Following the principle that the CH node can only select the lower-level CH node as RN, the multi-hop routing backbone network is established.

III. SYSTEM MODEL

A. NETWORK MODEL

Before describing the given algorithm in detail, the network model employed in the experiment is introduced. Some related network model assumptions are given per below:

- All sensor nodes are randomly deployed in the target region, and after the deployment phase, both the sensor nodes and the BSs are stationary.
- All sensor nodes know their location clearly after the deployment phase.
- All sensor nodes are able to change the transmission power according to the distance to the receiver nodes.
- The initial energy of all sensor nodes is the same.
- The processing power and energy supply of the BS are infinite.

B. ENERGY MODEL

The energy dissipation model in simulation employs the first order radio model [4]. Equation (1) represents the energy consumption in transmitting l bits of data to d distance.

$$E_{Tx}(l, d) = \begin{cases} lE_{elec} + l\epsilon_{fs}d^2 & d < d_0 \\ lE_{elec} + l\epsilon_{mp}d^4 & d \geq d_0 \end{cases} \tag{1}$$

Equation (2) represents the energy consumption in receiving l bits of data.

$$E_{Rx}(l, d) = lE_{elec} \tag{2}$$

Here, l is the number of transmitted information bits, d is the distance between the transmitter and the receiver, d_0 is the transmission distance threshold. E_{elec} indicates the energy consumed to run the transmitter or receiver circuitry. If the distance between the transmitter and the receiver is less than a threshold d_0, ϵ_{fs} indicates the energy consumed in the free space model; otherwise, ϵ_{mp} refers to the multi-path model is used. The value of d_0 is usually calculated as:

$$d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}} \tag{3}$$

Taking into account the correlation between the sensing data of nearby nodes, the CH nodes utilize data aggregation technology [33]–[35] to aggregate intra-cluster data to reduce data redundancy. Assuming global time synchronization, the data aggregation mechanism can effectively reduce network traffic, but it will increase the network communication delay. Data aggregation is divided into constant aggregation [14] and increasing aggregation [15] according to the size of the aggregated data packet. This paper utilizes the increasing aggregation (IA) model. The CH nodes form smaller data packet according to a certain aggregation ratio. The length of the aggregated data is calculated by the following equation:

$$L_{agg} = L_{rec} + L_{rec} \times \varepsilon \times N \tag{4}$$

where L_{agg} represents the length of the aggregated data packet, L_{rec} represents the length of the received data packet, $\varepsilon (0 \leq \varepsilon \leq 1)$ is the aggregation ratio and N is the number of CMs.

Energy consumption in data aggregation is represented by E_{DA}. In summary, the energy dissipation of the CH node is expressed by the following equation:

$$E_{CH} = E_{Rx} + E_{DA} + E_{Tx} \tag{5}$$

The energy dissipation of the non-CH node is calculated using Equation (6).

$$E_{non-CH} = E_{Tx} \tag{6}$$

IV. PROPOSED ALGORITHM

In this section, the proposed algorithm EAKDE is described in detail. EAKDE is a distributed unequal clustering routing algorithm. In order to determine the priority of a node to compete for CH, EAKDE is based on fuzzy logic which is taking both residual energy and distance to BS parameters into consideration. Moreover, EAKDE utilizes the adaptive kernel density estimation algorithm [37], [38] to assign appropriate cluster radius according to the local node conditions in each round. After the clustering phase, a multi-hop routing backbone network is established to perform data transmission. EAKDE consists of the following phases: cluster head election, cluster radius calculation, cluster formation and routing process.

A. CLUSTER HEAD ELECTION

After the sensor nodes are randomly deployed, each node can locate its own location [39]. The BS periodically broadcasts a *BS_ADV* message, including the base station id and the base station location. After receiving a *BS_ADV* message, node $S(i)$ stores the base station id and the base station location information, and calculates the distance to BS using Equation (7).

$$Dist_{BS}(S_i) = \sqrt{(X_i - X_{BS})^2 + (Y_i - Y_{BS})^2} \tag{7}$$

If a sensor node is closer to BS, it has higher ability to compete for CH and the greater probability of becoming a hot spot. In addition, it is obvious that the residual energy of the node is positively correlated with the ability to compete for CH. Therefore, EAKDE employs both residual energy and distance to BS to determine the priority of each

TABLE 1. Fuzzy if-then mapping rules for CH priority competition in EAKDE.

Rules	Residual energy	Distance to BS	CH priority
1	Low	Near	Rather weak
2	Low	Medium	Weak
3	Low	Far	Very Weak
4	Medium	Near	Medium Strong
5	Medium	Medium	Medium
6	Medium	Far	Medium Weak
7	High	Near	Very Strong
8	High	Medium	Strong
9	High	Far	Rather Strong

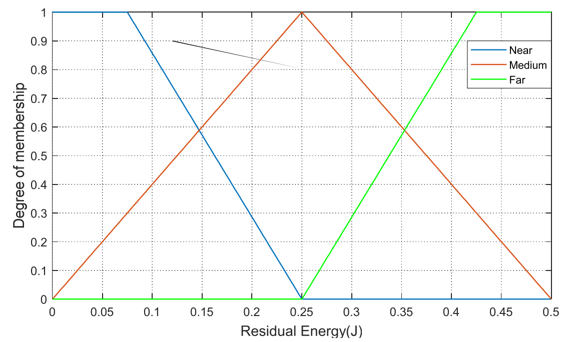

FIGURE 1. Fuzzy set - fuzzy input variable ***residual energy***.

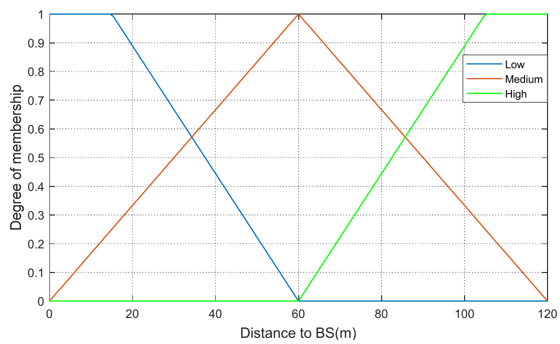

FIGURE 2. Fuzzy set - fuzzy input variable ***distance to BS***.

node to compete for CH. The CH competition uncertainty is handled by predefined fuzzy if-then mapping rules. These fuzzy if-then mapping rules are given in Table 1. The fuzzy inference technique uses the Mamdani method, and the center of area (COA) method is used for defuzzification of CH competition.

The ability of a node to act as CH, denoted as ‘chance’, changes dynamically, because EAKDE employs residual energy and distance to BS as fuzzy input variables. The fuzzy set of residual energy as input variable is demonstrated in Figure 1. The fuzzy linguistic variables for this fuzzy set are *Low*, *Medium* and *High*. The membership functions of *Low* and *High* are trapezoidal membership functions, and the trigonometric membership function is used for *Medium*.

Another fuzzy input variable is the distance to BS. The fuzzy set that describes the distance to BS input variable is depicted in Figure 2. *Near*, *Medium* and *Far* are the linguistic variables of this fuzzy set. *Near* and *Far* linguistic variables have a trapezoidal membership function, and *Medium* has a trigonometric membership function.

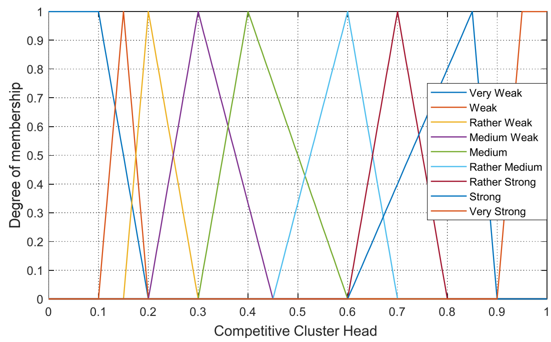

FIGURE 3. **Fuzzy set - fuzzy output variable 'chance'.**

The 'chance' as the only fuzzy output variable is denoted as u_i. The fuzzy set of 'chance' is illustrated in detail per Figure 3. There are 9 linguistic variables, i.e., *Very Weak*, *Weak*, *Rather Weak*, *Medium Weak*, *Medium*, *Medium Strong*, *Rather Strong*, *Strong*, *Very Strong*. *Very Weak* and *Very Strong* have the trapezoidal membership function, and the other linguistic variables have the trigonometric membership function.

There will be the strongest ability to act as CH for a node which has the most residual energy and is closest to BS. Conversely, the node with the least residual energy and the farthest from BS has the weakest CH competition ability. The ability of other nodes lies between these two extremes.

Each sensor node introduces a time delay T_i [32]. Once the timer of a node expires, it elects itself as CH and advertises a *CH_ADV* message. The specific calculation method of T_i is as follows:

$$T_i = \alpha \times (1 - u_i) \times T_C \tag{8}$$

where T_c is the maximum allowed waiting time. α, a random variable in [0.9,1], is mainly utilized to distinguish u_i, because there may be different nodes with the same 'chance'. Obviously, T_i decreases as u_i increases, meaning that the 'chance' of node $S(i)$ is greater.

B. CLUSTER RADIUS CALCULATION

The cluster radius is critical to the network lifetime. For calculating the cluster radius, mainly significant parameters are considered and described as follows:

1) NEIGHBOR NODE DENSITY

Assuming a local area, the sensor nodes are more densely distributed, then the cluster radius should be decreased to reduce the energy consumption of the CH node and avoid it failing too fast. Conversely, for a local area where the sensor nodes are sparsely distributed, the cluster radius can be appropriately expanded. A node transmits a *SN_ADV* message with a fixed radius, including the node id and node location. Once neighbor nodes receiving the broadcast message, they will return a *SN_ACK* message, including the residual energy, node id, and node location. Then this node stores each neighbor node information in a corresponding location in the neighbor list. The list is updated according to the current neighbors conditions at each round. The neighbor node density (node degree) is calculated by Equation (9), where N is the number of all sensor nodes in WSNs.

$$Density\ (i) = \frac{|Neighbor(i)|}{N} \tag{9}$$

2) NEIGHBOR NODE DISPERSION

The dispersion of the neighbors affects the size of cluster radius. It can be seen as another representation of local distance. The dispersion of the neighbors is computed by Equation (10).

$$Nei\_Disp\ (i) = \sum\nolimits_{j \in Neighbor(i)} exp(-Dist(S_i, S_j)^2/2\hat{\sigma}^2) \tag{10}$$

where $\hat{\sigma}$ is the standard deviation of the abscissa and ordinate values in the neighbors of node $S(i)$.

The energy dissipation of CMs for transmitting data packets is positively correlated with the distance to CH. This means that, the node distribution is relatively discrete, the CMs need to consume more energy to transmit data packets to CH. Therefore, the cluster radius should be appropriately decreased to reduce the intra-cluster energy dissipation. On the contrary, for the case where the node distribution is relatively concentrated, the cluster radius can be appropriately increased.

3) NODE RELATIVE RESIDUAL ENERGY

It is easy to understand that the more residual energy in the CH node, the larger the cluster radius should be. As network usage time passes, energy is continuously consumed, resulting in a decrease in the cluster radius. The relative residual energy, i.e., $Energy(i)$, is expressed as $\frac{Energy_{res}(i)}{Energy_{init}(i)}$. The nearby nodes generally have the same characteristics. So if a local area has more average relative residual energy, the area has a larger cluster radius.

4) RELATIVE DISTANCE TO BS

Considering the hot spot problem in WSNs, the CH nodes near to BS will bear higher data traffic, which leads to earlier death of CHs. Therefore, for the region closer to BS, decreasing cluster radius can effectively reduce the load of CHs. Assume that the sensor deployment region is a rectangular area. As shown in Figure 4, the point farthest from BS is the four vertices of the region, i.e., (0, 0), (X, 0), (0, Y), and (X, Y).

The relative distance to BS is calculated by Equation (11).

$$Dist\_BS(i) = \frac{Dist_{BS}\ (S_i)}{\max\ (d_1, d_2, d_3, d_4)} \tag{11}$$

According to the above four primary parameters, EAKDE calculates the adaptive cluster radius based on the local neighbor nodes conditions. The calculation of cluster radius is mainly divided into three steps.

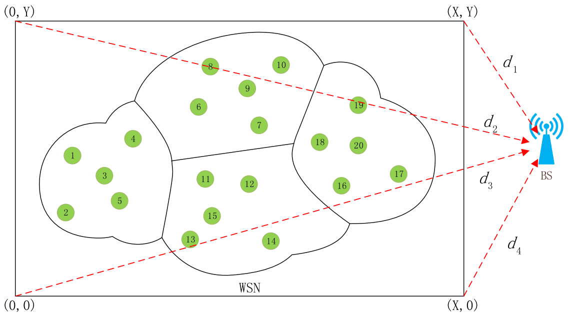

FIGURE 4. Sensor nodes and BS location architecture diagram.

Step 1 (Kernel density estimation): First, we give a kernel density estimate based on a fixed bandwidth under the global conditions. The pilot estimation $\hat{f}_H(l \mid E, D)$ of the distribution of sensor nodes on a position l is given by

$$\hat{f}_H(l \mid E, D) = \frac{1}{M}\sum_{i=1}^{N}\left[\frac{Density(i)}{Dist\_BS(i) \cdot Energy(i)} \cdot K_H(l - l_i)\right] \tag{12}$$

in which

$$Density(i) = \frac{|Neighbor(i)|}{N} \tag{13}$$

$$M = \sum_{i=1}^{N}\frac{Density(i)}{Dist\_BS(i) \cdot Energy(i)} \tag{14}$$

$$K_H(l - l_i) = \frac{1}{2\pi H_1 H_2}\exp\left(-\frac{(x - x_i)^2}{2H_1^2} - \frac{(y - y_i)^2}{2H_2^2}\right) \tag{15}$$

where $l_i (i = 1, 2 \cdots N)$ represents the geographic location of sensor nodes in WSNs, and $K_H(l - l_i)$ is a normal kernel function with a fixed bandwidth. The fixed bandwidth H consists of two global bandwidths (H_1, H_2). According to the mean integrated squared error minimization [38], the optimal (H_1, H_2) is given by

$$H_1 \approx 1.06 N^{-\frac{1}{5}}\hat{\sigma}_x \tag{16}$$

$$H_2 \approx 1.06 N^{-\frac{1}{5}}\hat{\sigma}_y \tag{17}$$

where the standard deviation of the abscissa of node $S(i)$ is given in (18), as shown at the bottom of this page, and the standard deviation of the ordinate of node $S(i)$ is given in (19), as shown at the bottom of this page.

Step 2 (Local bandwidth determination): The pilot estimation function $\hat{f}_H(l \mid E, D)$ is mainly affected by four parameters: $Dist\_BS(i)$, $Nei\_Disp(i)$, $Density(i)$ and $Energy(i)$. It can be found that when node $S(i)$ is far from the position to be estimated, the kernel function $K_H(l - l_i)$ is close to zero, and the contribution of this node to the pilot estimation function $\hat{f}_H(l \mid E, D)$ is almost zero. Therefore, it is very redundant to calculate the cumulative contribution of all nodes under global conditions. Considering the low processing power and small memory of sensor nodes, it is obviously not feasible to calculate $\hat{f}_H(l \mid E, D)$ under the global conditions. Therefore, the Nth-order nearest-neighbor of the location to be estimated is considered. While reducing the computational complexity, the space of neighbor list is saved. Here, the Nth-order nearest-neighbor of the position to be estimated is represented by a distance threshold CR (Cluster Radius).

Step 3 (Adaptive bandwidth determination: The neighbor nodes distribution of each node is different. Therefore, local bandwidth is obviously not applicable to calculate the cluster radius. In order to avoid local bandwidth defects, adaptive bandwidth determination h_i is used which is expressed with Equation (20).

$$h_i = g^{\gamma} \cdot \hat{f}_H(l_i \mid E, D)^{-\gamma} \tag{20}$$

where γ is a sensitive factor, with $0 \leq \gamma \leq 1$, and the larger γ indicates that h_i is more sensitive to the pilot estimation function $\hat{f}_H(l \mid E, D)$. g is the geometric mean, which means the geometric mean of h_i is equal to one.

$$g = \sqrt[n]{\prod_{i=1}^{n}\hat{f}_H(l_i|E, D)}, \quad n = |Neighbor(i)| \tag{21}$$

h_i is related to node location, node condition, and analysis scale. The smaller h_i is suitable for revealing the distribution of local node conditions, while the larger h_i can make the distribution of global node conditions more obvious. Different locations should adopt different analysis scales, and the adaptive bandwidth h_i can truly fit the distribution of local nodes. Therefore, the adaptive bandwidth h_i can be used to fit the cluster radius.

The BS sends the static parameters of Minimum Cluster Radius (CR_{min}) and Maximum Cluster Radius (CR_{max}) to all nodes in WSNs while advertising a BS_ADV message. $CR(i)$ denotes the cluster radius in Equation. (22).

$$CR(i) = h_i \cdot CR \tag{22}$$

$$\hat{\sigma}_x = \sqrt{\frac{1}{M}\sum_{i=1}^{N}\left(\frac{Density(i) \cdot x_i}{Dist\_BS(i) \cdot Energy(i)} - \frac{1}{M}\sum_{j=1}^{N}\frac{Density(j) \cdot x_j}{Dist\_BS(i) \cdot Energy(j)}\right)^2} \tag{18}$$

$$\hat{\sigma}_y = \sqrt{\frac{1}{M}\sum_{i=1}^{N}\left(\frac{Density(i) \cdot y_i}{Dist\_BS(i) \cdot Energy(i)} - \frac{1}{M}\sum_{j=1}^{N}\frac{Density(j) \cdot y_j}{Dist\_BS(i) \cdot Energy(j)}\right)^2} \tag{19}$$

Each node adjusts the size of the cluster radius according to the static parameters, i.e., $CR_{min} \leq CR(i) \leq CR_{max}$. It can be seen that $CR(i)$ is inversely proportional to the pilot estimation function $\hat{f}_H(l_i \mid E, D)$.

In summary, the effects of the four parameters on the adaptive cluster radius $CR(i)$ are as follows:

- The larger the distribution density of neighbor nodes, the smaller the cluster radius; on the contrary, the larger the cluster radius.
- If the distribution of neighbor nodes is discrete, the pilot estimate $\hat{f}_H(l_i \mid E, D)$ will be larger, and the cluster radius becomes very smaller. Conversely, for the case where the neighbor nodes are concentrated, the cluster radius is larger.
- As the average residual energy of neighbor nodes decreases, the cluster radius becomes smaller.
- The CH nodes near to BS will bear higher data traffic and consume their energy. Considering the load balance of hot spots, the cluster radius of them must be reduced.

The effect of above parameters on the decision to the adaptive cluster radius is demonstrated by the examples in Table 2. In these examples, the BS is placed at the center of region of internet, and the maximum distance to BS is 70.7m. In examples 1 and 2, as it approaches the BS, the cluster radius of the sensor node decreases. In examples 1 and 9, the relative distance to BS is identical, but energy levels and neighbor node density are different. The sensor node which has lower energy and higher neighbor node density has a smaller cluster radius.

TABLE 2. Sensor node's information.

Node	$Density(i)$	$Nei\_Disp(i)$	$Energy(i)$	$Dist\_BS(i)$	$CR(i)$
1	0.14	2.51e-04	0.83	25.85	15.18
2	0.13	2.57e-04	0.83	27.23	16.24
3	0.06	2.23e-05	0.82	58.19	26.75
4	0.13	2.04e-04	0.77	35.09	17.27
5	0.11	1.13e-04	0.74	40.57	21.31
6	0.04	7.76e-06	0.63	64.76	31.19
7	0.10	2.39e-04	0.58	43.52	19.58
8	0.05	1.89e-05	0.62	58.14	31.10
9	0.09	2.12e-04	0.34	28.71	8.44

C. CLUSTER FORMATION

When the timer T_i is reached, node $S(i)$ elects itself as CH and calculates its cluster radius according to step 3. A CH will broadcast a *CH_ADV* message within its cluster radius, including the CH id, CH location and CH energy. Once node $S(j)$ receives the message, it abandons the right to CH selection and immediately joins the cluster. If multiple *CH_ADV* messages are received, node $S(j)$ needs to weigh the energy consumption of joining each CH. Let $CH\_Cost(S_j, CH_i)$ denote the cost value. This cost function takes into account the following parameters.

1) Distance from CH to BS: The further the CH is from BS, it means that the CH needs to consume more energy to transmit data packets. It leads to

$$CH\_Cost(S_j, CH_i) \propto Dist_{BS}(CH_i) \tag{23}$$

2) Distance from non-CH to CH: A non-CH node can only transmit data packets to a CH node, so for a non-CH node, it is preferred to join the CH closest to it. It leads to

$$CH\_Cost(S_j, CH_i) \propto Dist(S_j, CH_i) \tag{24}$$

3) Direction of CH to BS: On the basis of parameters 1 and 2, a non-CH node is more inclined to join the CH towards BS. δ represents the angle between the $S(j)$ to CH_i connection and the $S(j)$ to BS connection. It leads to

$$CH\_Cost(S_j, CH_i) \propto \frac{1}{cos\delta} \tag{25}$$

4) Relative residual energy of CH: A non-CH node is biased towards joining a CH with higher relative residual energy. It leads to

$$CH\_Cost(S_j, CH_i) \propto \frac{1}{Energy(CH_i)} \tag{26}$$

In summary, the energy cost value of $S(j)$ to join each CH can be computed by the following equation.

$$CH\_Cost(S_j, CH_i) \propto \frac{Dist_{BS}(CH_i) \cdot Dist(S_j, CH_i)}{Energy\,(CH_i) \cdot (1 + cos\delta)} \tag{27}$$

A non-CH node calculates the cost of joining ith CH and joins the cluster with the lowest cost, i.e., $Min(CH\_Cost(S_j, CH_i))$. Then it sends a *CH_JOIN* message to the corresponding CH, including node id.

D. ROUTING PROCESS

CMs can only transmit data packets to their corresponding CH (not directly to BS). A CH node receives data packets transmitted by CMs and performs data aggregation. A multi-hop routing protocol is used between CHs to transmit data packets to BS. In order to minimize the routing path, the next hop of CH should be toward the direction of BS, thereby reducing the energy consumption in the routing transmission. Each CH calculates the routing level according to its distance to BS, i.e., $level(CH_i)$ (rounded up), is given by

$$level(CH_i) = \left| \frac{Dist_{BS}(CH_i)}{CR_{max}} \right| \tag{28}$$

where $level(CH_i) = 1, 2, \cdots k$. Obviously, a CH node with a higher level is farther away from BS. The CHs with routing level equal to 1 are directly routed to BS. A high-level CH node can only use a low-level CH node as a parent node (PN) and select one of them as a relay node (RN). The CHs broadcast the *RN_ADV* message within the range $k \times CR_{max}$ (initially $k = 2$). If the *RN_ACK* message sent by a lower-level CH is not received, the k value is increased until a *RN_ACK* message is received.

If multiple *RN_ACK* messages are received, CH_i needs to weigh the cost of sending data packets to each PN. $Cost(CH_{i,j})$ denotes the cost value, and it takes into account some parameters, including the transmitted energy consumption e^t_{ij}, the received energy consumption e^r_{ij}, and the residual energy of $CH_{i,j}$ [40]. e^t_{ij} and e^r_{ij} can be calculated according to Equations (1) and (2). Therefore, the cost of transferring data packets to the next hop node, is given by

$$Cost(CH_{i,j}) = e^t_{ij} \cdot Energy\,(CH_i)^{-\beta} + e^r_{ij} \cdot Energy\left(CH_j\right)^{\beta-1} \quad (29)$$

where $\beta (0 \leq \beta \leq 1)$ is the residual energy weight of $CH_{i,j}$. According to the cost estimate, a CH node selects the optimal RN among its PNs. The optimal choice is as follows:

$$RN\,(CH_i) = \min\left\{Cost\left(CH_{i,j}\right)\right\}, j \in \forall PN(CH_i) \quad (30)$$

Figure 5 illustrates the EAKDE multi-hop routing process for data packet transmission between CHs. For example, for node 3, node 9 and node 15 are chosen as the candidate RN. According to the multi-hop routing protocol, the optimal RN node 15 is selected. Node 9 and node 15 both utilize node 20 as a RN. The entire multi-hop routing backbone network in WSNs is shown in Figure 5.

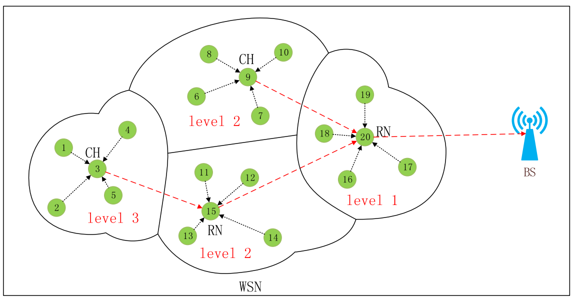

FIGURE 5. Multi-hop routing process in EAKDE.

V. EXPERIMENTS AND ANALYSIS

In this section, we use several experiments to evaluate the proposed algorithm EAKDE. We have performed simulation experiments in Matlab R2017b. Considering the impact of the BS location on the experimental results, two different network scenarios (i.e., Scenario 1 and Scenario 2) are used here. In Scenario 1, the BS is located at the corner of the region of interest (ROI), and in Scenario 2, the BS is placed at the center of ROI. The ROI is 100 m × 100 m, and the total number of sensor nodes in WSNs is 200. These scenarios are illustrated respectively in Figure 6 and Figure 7.

In the simulation experiment environment, the network runs in rounds. Each round is divided into four phases according to the proposed algorithm EAKDE: cluster head election, cluster radius calculation, cluster formation and routing process. The other clustering routing algorithm characteristics used for experimental comparisons in this paper are shown in Table 3, including LEACH, EAMR, imp-K-means, EAUCF, and DFCR.

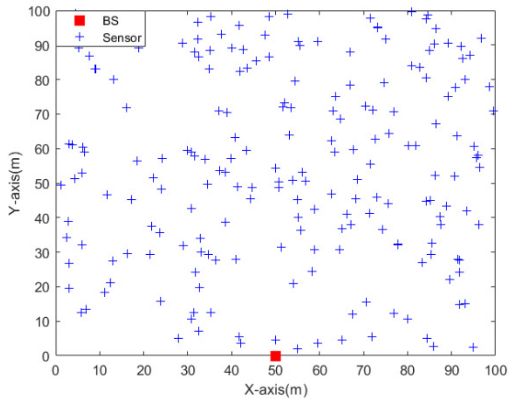

FIGURE 6. Scenario 1 - BS at the corner of ROI.

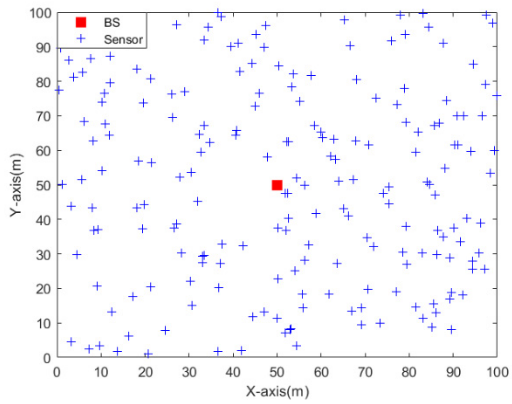

FIGURE 7. Scenario 2 - BS at the center of ROI.

According to Equation (31) [29], the desired percentage p of CHs for LEACH, EAMR, and EAUCF is set to 0.2, and the desired percentage p of CHs for imp-K-means is set to 0.05. The CH replacement threshold value of EAMR [20] is set to 0.04 in Scenario 1 and 0.02 in Scenario 2. Imp-K-means takes into account the residual energy and the distance to the centroid when elects the CHs, and the parameter weights are set to 0.8 and 0.2 in all of the scenarios respectively. The maximum distance to BS is 111.8m and the maximum competition radius [31] is set to 55m in Scenario 1 for EAUCF. In Scenario 2, the farthest node is 70.7m away from BS and the maximum competition radius is set to 35m for EAUCF. The maximum cluster radius of DFCR is set to 20m in all of the scenarios. For EAKDE, CR_{min} is set to 10 m, CR_{max} is set to 30 m, and the values of weight β and γ are set to 0.5 in all of the scenarios. Figure 8 and Figure 9 illustrate the resulting cluster layout by using the EAKDE algorithm.

$$p = \frac{\sqrt{n}}{\sqrt{2\pi}} \cdot \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}} \cdot \frac{\sqrt{A}}{(0.765 \times \sqrt{A} \times 0.5)^2} \cdot \frac{1}{n} \quad (31)$$

The experiment mainly evaluates the algorithm from two metrics of network lifetime and energy efficiency. The simulation parameters and their values are listed in Table 4.

TABLE 3. Comparison of the clustering approaches.

Protocol	Inter-cluster topology	Location awareness	Energy awareness	CH selection	Cluster count	Cluster size	Method
LEACH	Direct	Not Required	Not Required	Random	Variable	Equal	Distributed
EAMR	Multi-hop	Not Required	Required	Random	Constant	Equal	Distributed
imp-K-means	Direct	Required	Required	Deterministic	Variable	Equal	Centralized
EAUCF	Multi-hop	Required	Required	Deterministic	Variable	Variable	Distributed
DFCR	Multi-hop	Required	Required	Deterministic	Variable	Variable	Distributed

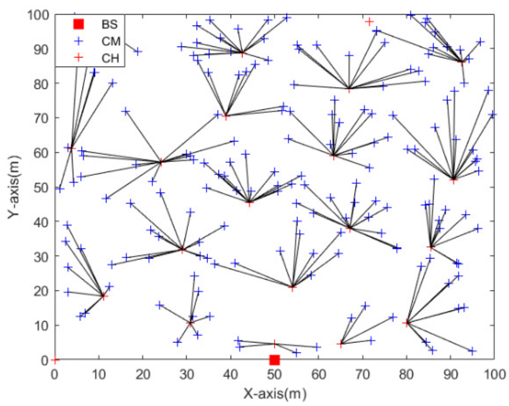

FIGURE 8. Scenario 1 - the resulting cluster layout.

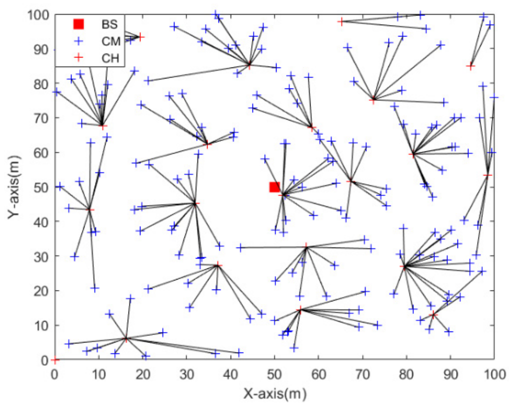

FIGURE 9. Scenario 2 - the resulting cluster layout.

TABLE 4. Simulation parameters and values.

Parameter name	Notation	Value
Target area	A	100 m × 100 m
BS location	BS	(0 m, 50 m)、(50 m, 50 m)
No. of sensors	N	200
Initial energy	E_0	0.5 J
Data aggregated energy	E_{DA}	5 nJ/bit
Transmitter/Receiver	E_{elec}	50 nJ/bit
Free space	ϵ_{fs}	10 pJ/bit/m^2
Multi-path space	ϵ_{mp}	0.0013 pJ/bit/m^4
Data packet size	D_p	500 bytes
Control packet size	C_p	25 bytes
Time delay	T_c	20 s
Aggregation ratio	ε	10%
Cluster Radius	CR	20 m

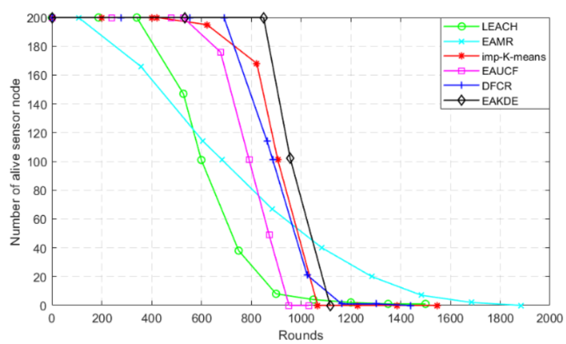

FIGURE 10. Scenario 1 - number of alive sensor nodes per round.

A. NETWORK LIFETIME

The most important concern in WSNs is the network lifetime. Handy *et al.* [36] used the first node dies (FND), half of the nodes alive (HNA) and last node dies (LND) to estimate the network lifetime. However, after more than half of sensor nodes die, the WSNs almost fails in most cases. Therefore, only FND and HNA metrics are chosen to evaluate the network lifetime.

In order to ensure the reliability of the results, every scenario is simulated 50 times, and the average values are taken. In each round of Scenarios 1 and 2, the distribution of the number of alive nodes for each algorithm is depicted in Figure 10 and Figure 11.

Figure 10 and Figure 11 clearly depict that EAKDE is more stable than other algorithms, because the FND metric for EAKDE begins later and node death continues linearly. Although for EAMR death of sensor nodes is earlier than other algorithms, it has the optimal LND performance. Since EAMR adopts fixed clusters mechanism, the overhead of forming new clusters at every round is saved, and the entire network lifetime is increased. However, in most applications, after more than half of nodes die, the WSNs is no longer considered valid. So for EAMR, the network lifetime is not regarded as optimal.

Imp-K-means is slightly lower than EAUCF and DFCR for the FND metric, but performs better for the HNA metric. Since EAUCF and DFCR employ the unequal clustering mechanism to optimize the cluster size, nodes become useless later than imp-K-means. Imp-K-means is just the opposite.

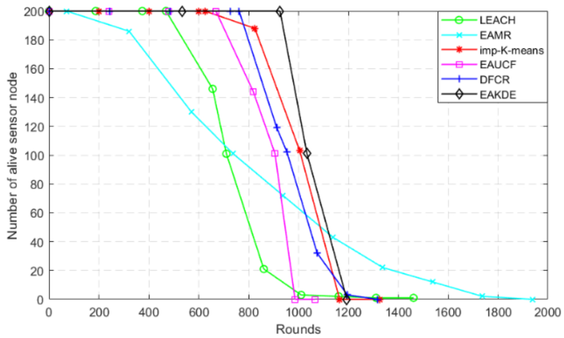

FIGURE 11. Scenario 2 - number of alive sensor nodes per round.

TABLE 5. FND and HNA of clustering algorithms in different scenarios.

Algorithm	Scenario 1		Scenario 2	
	FND	HNA	FND	HNA
LEACH	342	601	425	714
EAMR	99	685	82	740
imp-K-means	424	904	626	1006
EAUCF	545	799	675	912
DFCR	672	890	780	963
EAKDE	847	950	921	1025

It employs the equal clustering mechanism to balance the entire network load, so it has better performance for the HNA metric. However, imp-K-means is centralized and only applies to scenarios where the uniform distribution of nodes. It does not get good network scalability.

Note that EAKDE is not dominant if the LND metric is considered, because it has some overhead for exchanging control information. But on the other hand, EAKDE has the best FND and HNA metrics, which means it has the best performance in terms of network stability and network lifetime. EAKDE considers the distribution of local sensor nodes dynamic information, adaptively determines the cluster radius, and effectively equalizes the energy consumption of sensor nodes in WSNs.

In order to clearly analyze the results, FND and HNA of each algorithm in Scenarios 1 and 2 have been shown in Table 5.

As seen in Table 4, EAKDE performs significantly better than other algorithms (i.e., LEACH, EAMR, imp-K-means, EAUCF, and DFCR) considering the FND and HNA metrics in Scenarios 1 and 2. The FND metric reflects the network stability of WSNs to some extent. Although EAMR has an optimal LND , its FND is small. Since EAMR uses a fixed CH mechanism, the CH is rotated when the CH energy is lower than a given threshold, thereby reducing the overhead of forming new clusters, but at the same time resulting in the node to fail too fast. EAKDE are 147.6% and 116.7% more efficient than LEACH considering the FND metric in Scenarios 1 and 2, 99.7% and 47.1% more efficient than imp-K-means, 55.4% and 36.4% more efficient than EAUCF, 26.0% and 18.1% more efficient than DFCR.

The HNA metric can more accurately reflect the network lifetime. EAKDE outperforms all of the other algorithms considering the HNA metric. The HNA performance of EAKDE is 58.1% and 43.6% higher than LEACH in Scenarios 1 and 2, that of EAMR 38.7% and 38.5%, that of imp-K-means 5.1% and 1.9%, that of EAUCF 18.9% and 12.4%, and that of DFCR 6.7% and 6.4%. It can be observed that under the premise of ensuring the validity of WSNs, the number of active nodes in each round of EAKDE is more than LEACH, EAMR, imp-K-means, EAUCF and DFCR. Therefore, it can be concluded that EAKDE has better performance than other algorithms, in terms of network stability and network lifetime.

B. ENERGY EFFICIENCY

The energy of sensor nodes is one of the critical constraints of WSNs. The energy consumption rate of a node often depends on the routing protocol used. Low power consumption means that the entire network lifetime will be extended and the stability of the network will be enhanced. Figure 12 and Figure 13 show the average residual energy per round of all nodes in WSNs for EAKDE, and compare it with LEACH, EMAR, imp-K-means, EAUCF and DFCR.

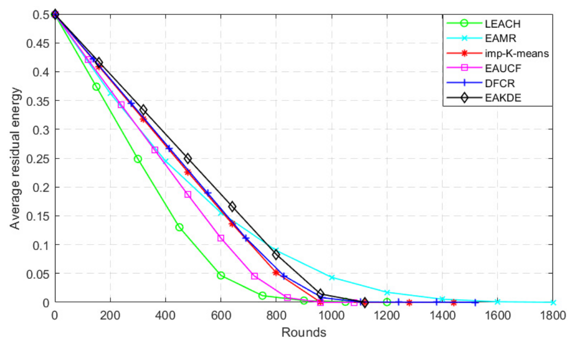

FIGURE 12. Scenario 1 - average residual energy per round.

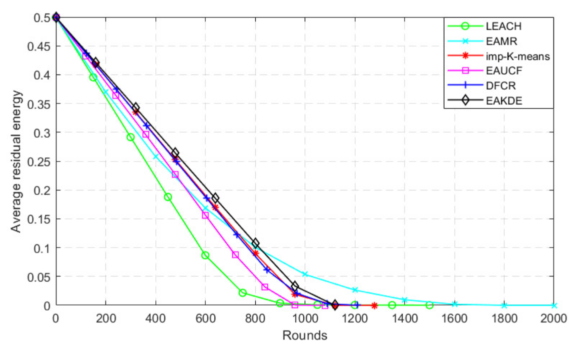

FIGURE 13. Scenario 2 - average residual energy per round.

As shown in Figure 12 and Figure 13, EAKDE is the most energy-efficient algorithm for Scenarios 1 and 2. Therefore, EAKDE has better network stability. For LEACH, EMAR, and EAUCF, because of the stochastic mechanism for CH selection, low energy nodes may be selected as CHs.

Imp-K-means utilizes the equal clustering mechanism, which does not effectively reduce energy consumption for non-uniformly distributed scenarios. In addition, for LEACH, EMAR and imp-K-means, the cluster formation strategy only takes into account the distance from non-CH to CH and the residual energy of CH, while ignoring the optimization of cluster size. Therefore, they can't effectively balance the energy dissipation of CHs.

EAUCF, DFCR and EAKDE all consider the dynamic changes in the network environment. Some node conditions are not constant, such as node failure or node movement (due to external factors like earthquake, storm and etc.), as well as the cluster radius of each round should change. Based on the density, dispersion, residual energy and relative distance to BS of the neighbors, EAKDE adaptively adjusts the cluster radius of each round, effectively balancing the CH load and reducing energy consumption. Therefore, EAKDE is more energy-efficient than other clustering algorithms.

The advantages of EAKDE are summarized as follows:

- For EAKDE, high energy nodes close to BS have the higher priority of acting as CHs, avoiding the defect that low energy nodes may be randomly selected as CHs.
- EAKDE is distributed in nature, and has better performance in network scalability and stability.
- Compared with other clustering routing algorithms, EAKDE does not need to set more threshold parameters or fuzzy if-then mapping rules, thus reducing the impact of human experience. EAKDE considers the dynamic change of node information, and adaptively determines the optimal cluster radius according to the network scenarios.
- In the cluster formation phase, non-CHs choose to join the nearest cluster, such as LEACH protocol. However, EAKDE takes into account the distance from CH to BS, the distance from non-CH to CH, the residual energy of CH and the direction of CH to BS, to optimize the composition of CMs.

VI. CONCLUSIONS AND FUTURE WORKS

A new unequal clustering algorithm EAKDE is proposed for WSNs in this paper, which aims to balance the workload among all sensor nodes. EAKDE consists of four main phases: cluster head election, cluster radius calculation, cluster formation and routing process. In order to adapt to the dynamic change of node conditions, EAKDE is committed to assigning the appropriate cluster radius to the sensor nodes by utilizing adaptive kernel density estimation algorithm. From the perspective of CH selection and cluster radius calculation, EAKDE avoids the effects of random uncertainty and human experience. It is distributed in nature and has a better performance compared to the other tested algorithms (i.e., LEACH, EMAR, imp-K-means, EAUCF and DFCR) in terms of the network lifetime and energy efficiency. These experimental results imply that EAKDE is a stable and energy-efficient unequal clustering algorithm for WSNs. In the future, we will consider using the adaptive kernel density estimation algorithm to elect the points-of-interest (POI) of the region. Let POI act as the active CH to further prolong the network lifetime.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[2] O. Powell, P. Leone, and J. Rolim, "Energy optimal data propagation in wireless sensor networks," *J. Parallel Distrib. Comput.*, vol. 67, no. 3, pp. 302–317, 2007.

[3] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Comput. Commun.*, vol. 30, nos. 14–15, pp. 2826–2841, Oct. 2007.

[4] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. IEEE 33rd Annu. Hawaii Int. Conf. Syst. Sci.*, Jan. 2000, p. 10.

[5] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.

[6] J. Xu, N. Jin, X. Lou, T. Peng, Q. Zhou, and Y. Chen, "Improvement of LEACH protocol for WSN," in *Proc. IEEE 9th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD)*, May 2012, pp. 2174–2177.

[7] D. Mahmood, N. Javaid, S. Mahmood, S. Qureshi, A. M. Memon, and T. Zaman, "MODLEACH: A Variant of LEACH for WSNs," in *Proc. IEEE 8th Int. Conf. Broadband Wireless Comput., Commun. Appl. (BWCCA)*, Oct. 2013, pp. 158–163.

[8] J.-L. Liu and C. V. Ravishankar, "LEACH-GA: Genetic algorithm-based energy-efficient adaptive clustering protocol for wireless sensor networks," *Int. J. Mach. Learn. Comput.*, vol. 1, no. 1, p. 79, 2011.

[9] M. Ye, C. Li, G. Chen, and J. Wu, "EECS: An energy efficient clustering scheme in wireless sensor networks," in *Proc. 24th IEEE Int. Perform., Comput., Commun. Conf. (IPCCC)*, Apr. 2005, pp. 535–540.

[10] R. Kumar and D. Kumar, "Hybrid swarm intelligence energy efficient clustered routing algorithm for wireless sensor networks," *J. Sensors*, vol. 2016, Aug. 2016, Art. no. 5836913.

[11] B. Singh and D. K. Lobiyal, "A novel energy-aware cluster head selection based on particle swarm optimization for wireless sensor networks," *Hum.-Centric Comput. Inf. Sci.*, vol. 2, no. 1, p. 13, 2012.

[12] S. Ghosh, S. Mondal, and U. Biswas, "A dominating set based modified LEACH using ant colony optimization for data gathering in WSN," in *Proc. 2nd Int. Conf. IEEE Adv. Elect., Electron., Inf., Commun. BioInformatics (AEEICB)*, Feb. 2016, pp. 390–396.

[13] D. C. Hoang, P. Yadav, R. Kumar, and S. K. Panda, "Real-time implementation of a harmony search algorithm-based clustering protocol for energy-efficient wireless sensor networks," *IEEE Trans Ind. Informat.*, vol. 10, no. 1, pp. 774–783, Feb. 2014.

[14] C. Joo and N. B. Shroff, "On the delay performance of in-network aggregation in lossy wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 22, no. 2, pp. 662–673, Apr. 2014.

[15] H. Teng, X. Liu, A. Liu, H. Shen, C. Huang, and T. Wang, "Adaptive transmission power control for reliable data forwarding in sensor based networks," *Wireless Commun. Mobile Comput.*, vol. 2018, Mar. 2018, Art. no. 2068375.

[16] L. Tan, Y. Gong, and G. Chen, "A balanced parallel clustering protocol for wireless sensor networks using k-means techniques," in *Proc. IEEE 2nd Int. Conf. Sensor Technol. Appl. (SENSORCOMM)*, Aug. 2008, pp. 300–305.

[17] A. Gachhadar and O. N. Acharya, "K-means based energy aware clustering algorithm in wireless sensor network," *Int. J. Sci. Eng. Res.*, vol. 5, no. 5, pp. 156–161, 2014.

[18] G. Y. Park, H. Kim, H. W. Jeong, and H. Y. Youn, "A novel cluster head selection method based on k-means algorithm for energy efficient wireless sensor network," in *Proc. 27th Int. Conf. IEEE Adv. Inf. Netw. Appl. Workshops (WAINA)*, Mar. 2013, pp. 910–915.

[19] M. Lehsaini and M. B. Benmahdi, "An improved k-means cluster-based routing scheme for wireless sensor networks," in *Proc. IEEE Int. Symp. Program. Syst. (ISPS)*, Apr. 2018, pp. 1–6.

[20] K. Cengiz and T. Dag, "Energy aware multi-hop routing protocol for WSNs," *IEEE Access*, vol. 6, pp. 2622–2633, 2018.

[21] W. Peizhuang, "Pattern recognition with fuzzy objective function algorithms," *SIAM Rev.*, vol. 25, no. 3, p. 442, 1983.
[22] D. C. Hoang, R. Kumar, and S. K. Panda, "Fuzzy c-means clustering protocol for wireless sensor networks," in *Proc. IEEE Int. Symp. Ind. Electron. (ISIE)*, Jul. 2010, pp. 3477–3482.
[23] R. Dutta, S. Gupta, and M. K. Das, "Low-energy adaptive unequal clustering protocol using fuzzy c-means in wireless sensor networks," *Wireless Pers. Commun.*, vol. 79, no. 2, pp. 1187–1209, 2014.
[24] J. S. Lee and W. L. Cheng, "Fuzzy-logic-based clustering approach for wireless sensor networks using energy predication," *IEEE Sensors J.*, vol. 12, no. 9, pp. 2891–2897, Sep. 2012.
[25] S. K. Das, S. Tripathi, and A. P. Burnwal, "Fuzzy based energy efficient multicast routing for ad-hoc network," in *Proc. IEEE 3rd Int. Conf. Comput., Commun., Control Inf. Technol. (C3IT)*, Feb. 2015, pp. 1–5.
[26] D. T. Hai, L. H. Son, and T. Le Vinh, "Novel fuzzy clustering scheme for 3D wireless sensor networks," *Appl. Soft Comput.*, vol. 54, pp. 141–149, May 2017.
[27] R. Logambigai and A. Kannan, "Fuzzy logic based unequal clustering for wireless sensor networks," *Wireless Netw.*, vol. 22, no. 3, pp. 945–957, Apr. 2016.
[28] N. T. Tam and D. T. Hai, "Improving lifetime and network connections of 3D wireless sensor networks based on fuzzy clustering and particle swarm optimization," *Wireless Netw.*, vol. 24, no. 5, pp. 1477–1490, 2018.
[29] J.-M. Kim, S.-H. Park, Y.-J. Han, and T.-M. Chung, "CHEF: Cluster head election mechanism using fuzzy logic in wireless sensor networks," in *Proc. IEEE 10th Int. Conf. Adv. Commun. Technol. (ICACT)*, vol. 1, Feb. 2008, pp. 654–659.
[30] B. Baranidharan and B. Santhi, "DUCF: Distributed load balancing unequal clustering in wireless sensor networks using fuzzy approach," *Appl. Soft Comput.*, vol. 40, pp. 495–506, Mar. 2016.
[31] H. Bagci and A. Yazici, "An energy aware fuzzy approach to unequal clustering in wireless sensor networks," *Appl. Soft Comput.*, vol. 13, no. 4, pp. 1741–1749, Apr. 2013.
[32] N. Mazumdar and H. Om, "Distributed fuzzy approach to unequal clustering and routing algorithm for wireless sensor networks," *Int. J. Commun. Syst.*, vol. 31, no. 12, p. e3709, 2018.
[33] Z. Wang, Q. Cao, H. Qi, H. Chen, and Q. Wang, "Cost-effective barrier coverage formation in heterogeneous wireless sensor networks," *Ad Hoc Netw.*, vol. 64, no. 9, pp. 65–79, 2017.
[34] Z. Wang, J. Liao, Q. Cao, H. Qi, and Z. Wang, "Achieving k-barrier coverage in hybrid directional sensor networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 7, pp. 1443–1455, Jul. 2014.
[35] M. Wu, Y. Wu, X. Liu, M. Ma, A. Liu, and M. Zhao, "Learning-based synchronous approach from forwarding nodes to reduce the delay for Industrial Internet of Things," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 10, 2018.
[36] M. J. Handy, M. Haase, and D. Timmermann, "Low energy adaptive clustering hierarchy with deterministic cluster-head selection," in *Proc. IEEE 4th Int. Workshop Mobile Wireless Commun. Netw.*, Sep. 2002, pp. 368–372.
[37] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Abingdon, U.K.: Routledge, 2018.
[38] J.-D. Zhang and C.-Y. Chow, "GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2015, pp. 443–452.
[39] J.-H. Chang and L. Tassiulas, "Maximum lifetime routing in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 12, no. 4, pp. 609–619, Aug. 2004.
[40] M. M. Afsar, and M.-H. Tayarani-N, "Clustering in sensor networks: A literature survey," *J. Netw. Comput. Appl.*, vol. 46, pp. 198–226, Nov. 2014.
[41] M. Singh and S. K. Soni, "A comprehensive review of fuzzy-based clustering techniques in wireless sensor networks," *Sensor Rev.*, vol. 37, no. 3, pp. 289–304, 2017.

FAGUI LIU received the M.S. degree from Beihang University, in 1991, and the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 2006, where she is currently a Professor with the School of Computer Science and Engineering. Her research interests include service computing, the Internet of Things, cloud computing, and big data.

YUFEI CHANG is currently pursuing the M.S. degree with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include fuzzy systems, wireless communications, and the Internet of Things.